

V. CONTRIBUTED PAPERS IN ARCHÆOLOGY

A. The Application of Matrix Index Systems to Archæological Materials

ROGER GREEN

INTRODUCTION

Data-processing and information retrieval systems are now familiar to the world of business and of science. They are being used more and more, but not in even manner. Many disciplines still fail to take advantage of them, not because these systems are unable to handle the relevant materials, but simply because it requires of the scientist a re-orientation in his existing *modus operandi*.

The best theoretical and practical case for the utilization of these systems in archæology was made in 1955 by Jean-Claude Gardin. He showed that with the enormous growth of documentation, the solution of better circulation of the information, within the group engaged in archæological research, lay in the adoption of these systems. This may take one of two major lines. The first is the addition of new outlets to an already existing network. For Pacific archæologists, the relevant COWA surveys, or the summaries provided by *Asian Perspectives*, are recent additions falling within this category. The alternative is a transformation of the rule of use, or the *code*, so as to increase the flow of information passing within the given network during any interval of time. To date, Pacific archæologists have been content to develop further the existing means of communication; but it is to this second alternative, the altering of the code by means of data-processing techniques that I wish to call your attention.

THE PROBLEM

In Pacific archæology, as in any selected area of research, we are to a large degree confronted with very much the same basic documentation. While our interests, and consequently our emphasis, vary considerably, we consult the same bulletins, journals, and other reference works; and study the same body of material objects recovered by excavation, be they fishhooks, axes, adzes, pottery, weapons, house types, or midden specimens, etc. As the complexity and quantity of the material increases, the tendency is to specialize on one or more given aspects of the total problem, frequently within a given sub-area. But if the results of our individual research is not to be of limited value, it has to be projected on a wider canvas. This will mean drawing on a wider field of documentation whereby we can make broader generalizations about the prehistory of the Pacific area as a whole, and thus link up our materials with other works in more embracing and significant hypotheses.

In the process of moving from his own material to that of others, each scholar develops, according to his interests, a catalogue of items that he deems essential to his own research, and with it, a working bibliography or bibliographies documenting

these and other items. But the procedures he adopts to file this material, or the principles by which he sorts out from a vast bulk of documentation, the data he seeks, is achieved in anything but a *uniform* manner.

More distressing, however, is the overlap and duplication in time, effort, and ability that goes into the construction of these private reference files. Consider for instance the number of people who have, at various times, assembled files on collections of adzes from the Pacific area. Consider also the frequency with which people have recourse to the same body of literature concerning these implements in order to document some particular point or points. Even with the best of the old style filing systems and indexes, the search may take considerable time.

To assist in the search, there are, of course, published bibliographies, indexes, and public catalogues, as well as organized storage repositories like museums. But all suffer from a common difficulty in that the data they contain exists in a bound form, to which easy access is not always possible. As to indexes and catalogues not only their production demands considerable effort, but also their publication is expensive and entails delay. The difficulties involved in the study of private and public collections are too well known to require enumeration. It is simply not possible to circulate the objects or information on them to every scholar who might find it useful.

The fact that these materials are in both cases organized along lines which do not readily permit regrouping is even more troublesome. That is, they are normally organized into hierarchical, or pyramidal systems in which objects are grouped by major categories, sub-categories, and sub-sub-categories which differ from book to book, index to index and museum collection to museum collection. If these categories are not pertinent to the research in question, or if the materials under consideration are distributed throughout various sub-sub-categories, the assembling of the basic documentation for any piece of research becomes a considerable task.

THE PRINCIPLES

The ability to regroup, repeatedly, the material being searched or analysed is one of the fundamental operations in research. The second is the ability to continually add new information to that body of material without necessitating its complete reorganization. In so far as the documents are objects, and thus open to repeated observation by all observers, it should be possible for various investigators to agree on the component elements which they exhibit, even if they never agree as to the significance which a given investigator wishes to attach to a particular element. Indeed, in existing archaeological classifications, there must be and is a large measure of agreement on a minimum number of features that should be recorded, regardless of how the investigator feels about their relative significance. The crucial point is the development of a means whereby an investigator can record these features, and all others he feels important, in a manner that avoids the necessity of ranking them in order of importance. This makes it possible for others to regroup the same data along other lines if they consider them more significant for their purpose.

A principle, in historical and structural research, demands that the documentation be arranged in such a manner that the operation of extracting the required information be both efficient and easy. Furthermore it demands that the documentation be open to additions by all investigators at any time with a minimum of effort. For this reason public catalogues, bibliographic indexes, and analytical files of objects must be of a type that are open-ended and capable of orientation along any set of lines chosen by the particular investigator. Hence any change in the existing code for archæological materials requires that the proposed apparatus be of a form which facilitates not only the old operations, but permits the material to be submitted to new investigations (Gardin 1955)—otherwise, we frustrate the very goal we seek by concentrating our energies on new means to old ends.

The solution Gardin has suggested is that scholars, in common or restricted fields of research (as for instance, those working in Pacific archæology, in which there is already an existing network of communication), construct an objective and public documentation with the data arranged and stored on file cards. The information recorded should be done in such a way that it may be easily and universally read; that is, the elements being described must be strictly enough defined as to their meaning, and thus reduced to a code which can be rapidly read into whatever language the investigator normally uses. Equally important is the adoption of a procedure or system whereby the file of documents can be rapidly cross-indexed and re-sorted. Finally all the cards in the system must be capable of rapid reproduction and the complete system low in cost.

SOLUTIONS

In France the creation of the CENTRE D'ANALYSE DOCUMENTAIRE POUR L'ARCHÉOLOGIE is the outcome of Gardin's proposals and an attempt to meet the conditions outlined above (Gardin 1959). This centre is presently designed to handle the documentation of Bronze Age weapons, tools and pottery; the iconography on items like classical Cretan pottery, cylinder seals, and Greek painted vases; and indexes for texts, on clay tablets and in the Koran. For its operations, the only equipment possessed by the Centre is that used for punching, duplicating, and storing the various files of cards. The key to the whole operation lies in the development of adequate codes for documentary purposes, and the use of a matrix-index system of information retrieval, called *Selecto*, which meets the various requirements of open-endedness, ease of operation, low cost, and ability to reorganize data along any lines.

In Britain, one of the first applications of a similar matrix-index system to archæological material was undertaken by Dark (1957) in preparing a system of documentation for all objects of the famous Benin art style.

In America, the best known application of punched cards to archæological materials is that connected with radiocarbon dates (Oswalt 1959, Radiocarbon Assn. 1959). This system used the more familiar edge-punched cards. In the Pacific, the one such use of edge-punched cards known to me is that presently in use at the Bishop Museum for the storage and analysis of data pertaining to more than 4,000 fishhooks from Hawaii (Emory *et al.* 1959).

These are but a few of a growing number of instances where punched cards and the use of data processing and information retrieval techniques have been applied to archæological materials with advantage. Perhaps the prospects were best summed up in the announcement in the NOTES AND NEWS section of *Antiquity* (1958) concerning the availability of a set of 3,000 punched cards for radiocarbon dates. It ended with the terse comment 'When we all have our cards and needles, a silent revolution in prehistory will have taken place'.

SYSTEMS

The foregoing statement is probably not unjustifiably optimistic, but the means to achieve it is not so clearly defined. There are a number of systems from which to choose, all of which have certain advantages and disadvantages depending on the material to be analysed. Some have certain advantages over others in their application to archæological materials and in the amount of time, money and effort which may reasonably be expended.

In the first set of systems the data is either punched into or written onto a single card. This becomes the complete document card for that item. The edge-punched cards used in the radiocarbon card file, or in the analysis of fishhooks at the Bishop Museum are examples of this form of system. Another is the 'système Dequeker' which the French originally used at the Institute of Archæology at Beyrouth. Here the punching covers the whole lower portion of the card. A third is that of the IBM type, in which the whole face of the card is both punched and sorted by machine, a system which is now being applied to the analysis of Peruvian fabrics by Junius Bird at the American Museum of Natural History.

All of these systems have certain disadvantages which militate against their widespread adoption in the analysis of the large collections of complex archæological materials. Systems requiring machine punching and sorting are beyond the means of most archæologists or institutions. The edge-punched card systems of the type used in the fishhooks analysis, while inexpensive, have proved to have insufficient categories available for the addition of new data, hence they will have to be re-punched on new cards designed to store more information or transferred to a completely open-ended system. Finally the sorting of edge-punched cards becomes a slow and laborious process when thousands of cards and multiple categories are entailed.

When the French set up their centre for documentation, they even found the more sophisticated system of Dequeker unsatisfactory for the purpose they now envisioned. First, the card capacity was too small. Second, if the documentation approached several tens of thousands, the method of searching through lots of 400 cards would be slow and time consuming, and demand many repeated manipulations. Lastly, as the collections enlarged, filing cabinets begin to take on cumbersome dimensions; their cost moreover reduces the chances for wide adoption. For these reasons, the French turned to another system (Gardin 1959: 15-17).

The second set of systems, sometimes called 'inverted punch card' or 'peek-a-boo' systems, are generally referred to matrix-index systems of information retrieval. They are characterized by the use of a set of index cards in which each

card represents a specific aspect or feature. In this system any object or specimen has the same position or number throughout the set of punched cards.

These systems also come in several varieties. In general, it may be said that they are: *i.* less expensive, especially the simpler among them; *ii.* open-ended, so that new sorting categories and documents may be continually added without upsetting the previous work; and *iii.* almost instantaneous in their selection of the documents containing any particular combination of elements that are desired. Even the more elaborate machine models are not beyond the budget of a museum or anthropology department, while the simpler systems are well within the range of do-it-yourself exponents.

I have described elsewhere the application of the simplest of these systems, called *Uniterm*, to small working bibliographies, site surveys, or artifact collections (Green 1961 *b*). A more recent application of this method to a file of radiocarbon dates from Oceania is in use at the Departments of Anthropology, in the Bernice P. Bishop Museum and the University of Auckland.

Basically the system consists of a small set of cards that permit cross-indexing the larger file of cards, each containing all the relevant information about any one date. New categories for sorting or new dates indexed under the old categories may be added at any time in a matter of minutes. More important, anyone with a list of the dates he considers relevant for his purpose and a knowledge of the principles involved can make his own set in a matter of several days, with little expenditure other than of time and the cost of several hundred plain index cards.

The next step beyond this requires the use of hand-punched cards. Each object is first assigned a single position on a card, for example row eight, column five, which can also be read as eighty-five. When the positions for all objects are assigned, those which exhibit the particular characteristic or feature in question are punched. The same process is applied to other cards for different characteristics or features exhibited by the objects in question. Because the object maintains an identical position on each card, whether it has that characteristic or not can be determined by whether or not that position on the card is punched. Objects with a given combination of features are identified by aligning the cards over a light source and identifying those positions at which the light shines through. In this system instead of all the data being punched on the same card, it is scattered among a number of punched cards. This feature is not a serious disadvantage, because all the data, including references, photo-file numbers, and a complete code for every characteristic exhibited by any single item are listed on a separate card filed numerically according to its position number. Thus you have two sets of cards, one a large file in numbered order, each card containing all the relevant information about that item; the other a far smaller set of punched cards which you can combine in endless fashion to obtain the position numbers for those items with the desired features. So if you want an adze with a curved back, modified butt, triangular section at the shoulder, and projection at the poll, you need only combine the four cards punched for these characteristics. Those positions, where the light shines through, indicated the numbers of the relevant adzes. It is this optical principle, which has lead some to refer to the system by the name 'peek-a-boo'.

Enlarged versions of these matrices, handling up to 10,000 specimens at one time may be used in the analysis of adzes, fishhooks, pottery, bibliography, or any other data that can be stored in a numbered series or file. This method of cross-indexing and searching documents for the required information, meets the basic requirements set out above for ease of operation, simplicity, and low cost, as well as providing the means for sorting by any set of criteria deemed relevant. Also the document or object cards which contain all the necessary reference material and all of the coded data can be easily reproduced and stored. There is some disadvantage, however, in the need for two sets of cards, although the second or index set, need never be large, and the first set is of the same size as in the other system.

With a file of standard object cards, bearing a printed code, any investigator is free to adopt whatever equipment he chooses, and proceed to index them. He need punch only for those aspects in which he is interested or for which he wishes to sort, thus fashioning the index to his own tastes. Thus the need for agreement here, lies, not in the necessity for adopting a specific variety of system or type of card for the storage and retrieval of this information, but merely in agreement to use matrix-index principles in the analysis of our basic documentation, and to arrive at a standard code for disseminating that information. It is not necessary to distribute punched cards, but rather to distribute object cards with coded information.

Of these two problems the more serious is the adoption of sufficiently uniform methods for coding the objects already in published sources, or contained in museums, or those currently being recovered from excavations. Processing them for use in a matrix-index system requires the adoption of lineal rather than pyramidal methods of classification, and the reduction of this 'classification' (description) to a code which is unambiguous in its meaning.

CODES

The production of such a code, unfortunately, is not simply a matter of agreement that can be reached around the conference table. It is a result of repeated preliminary attempts to define and classify ever larger bodies of the same materials, until the code covers precisely all the variations presented by the objects belonging to a class. It is then necessary to divide the variations into a number of subjects or categories, listing exhaustively the features and combinations possible under any sub-set, so that the particular features exhibited by any item drawn from that class have only one possible listing under each category. Four codes, and the theory and the principles involved in developing them, are clearly set forth by Gardin (1958).

A code which my wife and I developed from Gardin's work is now in use on Polynesian adzes. It has now undergone at least three major revisions. A more recent code for Polynesian fishhooks has already undergone at least two revisions. It is derived from an earlier one presently in use by the Bishop Museum. Before any of them achieve standards sufficiently rigorous for general use, they must be tested by a number of people, revised, and standardized to the point where a number of investigators can independently achieve reliable results on the same body of material. Furthermore, the distinctions we have made must meet with

widespread acceptance, and certain conventions agreed to by all who use the code, for it is absolutely necessary that certain arbitrary conventions be adopted at some points.

An excellent start has been made at the Bishop Museum with the fishhook material from Hawaii. Results from Mangareva and Tahiti, and examination of fishhooks from Polynesia in general, indicate that an expansion of their code to include the remainder of the Pacific material will demand certain major revisions (Green 1961 *a*). With the addition of other areas like New Zealand, or Japan, further revisions will doubtless be necessary. Nevertheless, the stage is set for discussion in the immediate future, of proposals for minimum codes and terms for Polynesian adzes and fishhooks. Work by other specialists on classes like food pounders, pottery, ornaments, weapons, and stone figurines, etc. could easily result in codes for these items as well. This leaves then, the question of whether an expenditure of time and effort in this direction is desirable.

CONCLUSION

As indicated at the beginning of this paper, I accept the position of Gardin that only a change in the *code* within our already existing network of communication, and not further additions to it, will allow us to solve our major problem. The size of the problem now is small compared with areas like Southwestern United States, Europe, or the Middle East; but the expansion of research in Pacific archæology will shortly begin to multiple the basic documentation beyond our means to master it fully. However, as was stressed in that discussion, a new apparatus must allow us not only to continue the old operations, but to engage in significant new research. The apparatus proposed here, that is the adoption of the principles of matrix index systems for the searching and analysis of materials stored on coded file cards, would be, I believe, a stimulus to new research. To implement such a system we now need the production of codes for various classes of objects which are the result of intensive preliminary studies. Beyond that there is a need for a committee and centre to test and standardize these codes, circulate them to scholars, and distribute periodically to those who wish, duplicate object cards containing information on the various items, as these are sent in by investigators. The basis for this we have in the Pacific Area Archæological Program and resolutions concerning it adopted at the Tenth Pacific Science Congress (Green 1961 *c*).

As to the question of whether the adoption of these procedures serves as an aid to analysis, one may cite the results achieved by their use on Hawaiian fishhooks (Emory *et al.* 1959). I can report the similar results from the application to the fishhook material from Mangareva. Finally, there are the first of a series of results from a study of 293 adzes from Central Eastern Polynesia in New Zealand Museums which were coded and measured. An analysis of this sample shows clearly both in graphical and statistical terms that whereas both the width and thickness at the shoulder or central portion of the adze is related directly to the length regardless of geographical location, nature of the material, or type of the adze, the ratio between width and thickness exhibits differences with respect to island groups, the geometric shape of the cross-section and a number of other features (Green and

Purcell 1961). Thus for the first time the nature of the relationship between three measurements on a large collection of adzes is explored and defined precisely; and the nature of the association between the most significant of these relationships, called the shoulder-index, and other features used to classify adzes, is examined in detail. Further new research of this type may be the expected result of adopting these proposals.

Finally, and of *equal importance*, the ease with which the old operations may continue to be performed, will be vastly improved.

BIBLIOGRAPHY

DARK, P. J. C.

- 1957 A preliminary catalogue of Benin art and technology: Some problems of material culture analysis, *JRAI*, **87**, 175-189.

EMORY, K. P., BONK, W. J. and SINOTO, Y. H.

- 1959 Hawaiian archæology: Fishhooks, *B.P. Bishop Museum Special Pub.* **47**, Honolulu.

GARDIN, JEAN-CLAUDE

- 1955 Problems of documentation, *Diogenes*, **11**, 85-101.
 1958 Four codes for the description of artifacts: an essay in archæological technique and theory, *AA*, **60**, 335-357.
 1959 *Centre d'Analyse Documentaire pour l'Archéologie*, Centre National de la Recherche Scientifique, Paris.

GREEN, R. C.

- 1961a Review of Emory, Bonk, and Sinoto, 'Hawaiian Archæology: Fishhooks' (*B.P. Bishop Mus. Pub.* **47**) in *JPS*, **70**, 139-144.
 1961b A simple matrix-index for working bibliographies, site surveys or small artefact collections, *JPS*, **70**, 239-243.
 1961c Tenth Pacific Science Congress: (1) Archæology in 'Pacific Commentary', *JPS*, **70**, 477-481.

GREEN, R. C. and PURCELL, D. W.

- 1961 The relationship of length, width, and thickness in Central Polynesian adzes, *JPS*, **70**, 451-465.

NOTES and NEWS

- 1958 Radio-carbon dating and archæology, *Antiquity*, **32**, 193-194.

OSWALT, W. H.

- 1957 A punched card system for radiocarbon dating, *American Antiquity*, **23**, 183-185.

RADIOCARBON DATES ASSOCIATION, INC.

- 1959 *A Manual for the Utilization of Edge-Punched Cards in the Radiocarbon File*, Andover Mass.